# CRAWLER TERFOKUS BERBASIS ONTOLOGI DENGAN MENGGUNAKAN GAUSSIAN-BASED ADAPTIVE THRESHOLDING SETTING (GATS)

**TIAZ ISKANDAR MUDA**
(Pembimbing : Heru Agus Santoso, Ph.D)
*Teknik Informatika - S1, FIK, Universitas Dian Nuswantoro*
*www.dinus.ac.id*
*Email : 111201307468@mhs.dinus.ac.id*

## ABSTRAK

E-Government Development Index (EGDI) pada tahun 2016 menyatakan bahwa Indonesia memiliki peringkat terendah dibandingkan dengan sebagian besar negara-negara di Asia Tenggara. Salah satu penyebabnya adalah tidak terpeliharanya situs E-government kerena belum optimalnya integrasi data/informasi. Agar dapat mengatasi masalah tersebut maka diusulkan pendekatan agregasi secara otomatis dan integrasi data/informasi pada situs E-government. Pendekatan ini dapat diimplementasikan untuk meningkatkan kapasitas situs E-government, misalnya untuk memelihara konten secara otomatis dan benchmarking situs pemerintah. Data atau informasi yang digunakan diperoleh dari 10 situs pemerintah daerah di Jawa Tengah dengan teknik crawling terfokus berbasis ontology. Dari proses crawling akan dihitung tingkat relevansinya dengan ontology berdasarkan similarity, kemudian akan dilakukan pengecekan terhadap ambang batas kemiripan data atau informasi yang disebut dengan thresholding. Pada penelitian ini digunakan Gaussian-based adaptive thresholding setting (GATS) untuk mengolah data thresholding terhadap relevansi data crawling. Percobaan crawling pada 10 website Â¬e-government pemerintahan daerah di Jawa Tengah berdasarkan pada 10 domain ontology yang ada, diperoleh 410 data relevan dari total 722 data crawling dimana diperoleh nilai total harvest-rate sebesar 56.79% dari 10 website e-government.

Kata Kunci : Crawling terfokus, ontology, Gaussian-based Adaptive Thresholding, e-government

# FOCUSED CRAWLER BASED ON ONTOLOGY USING GAUSSIAN-BASED ADAPTIVE THRESHOLDING SETTING (GATS)

**TIAZ ISKANDAR MUDA**
(Lecturer : Heru Agus Santoso, Ph.D)
*Bachelor of Informatics Engineering - S1, Faculty of Computer Science, DINUS University*
*www.dinus.ac.id*
*Email : 111201307468@mhs.dinus.ac.id*

## ABSTRACT

E-Government Development Index (EGDI) in 2016 stated that Indonesia has the lowest rating compared to most countries in Southeast Asia. One possible cause is not maintaining the E-government because they have not been optimal integration of data / information. In order to overcome these problems, the proposed approach automatically aggregation and integration of data / information on the E-government. This approach can be implemented to increase the capacity of the E-government, for example, to maintain content automatically and benchmarking government sites. Data or information used were obtained from 10 sites local governments in Central Java with a focused crawling technique based ontology. Crawling will be calculated level of relevance to the ontology based on similarity, then will be checked against the threshold of similarity data or information referred to thresholding. In this research used a Gaussian-based adaptive thresholding setting (GATS) for processing thresholding to the relevance of the data crawling. Experiments crawling on the 10 e-government website of local government in Central Java domain ontology based on the existing 10, 410 obtained relevant data from a total of 722 data is crawling which gained a total value of harvest-rate amounted to 56.79% of the 10 e-government website.

Keyword : Focused crawling, ontology, Gaussian-based Adaptive Thresholding, e-government